Distribution analysis of vitamin D highlights differences in population subgroups: preliminary observations from a pilot study in UK adults

B R Pal, T Marshall[1], C James and **N J Shaw[2]**

Department of Biochemistry, City Hospital, Birmingham, UK

[1]Department of Public Health and Epidemiology, University of Birmingham, Birmingham, UK

[2]Department of Endocrinology, Birmingham Children's Hospital, Birmingham, UK

(Requests for offprints should be addressed to N Shaw, Department of Endocrinology, Birmingham Children's Hospital, Steelhouse Lane, Birmingham B4 6NH, UK; Email: nick.shaw@bch.nhs.uk)

Abstract

There is no consensus between authors on the definition of a replete or deficient vitamin D state. Our aim was to describe a suitable method that could be used to compare vitamin D data in subject groups with small or large numbers.

Two hundred and forty indigenous asymptomatic, non-pregnant adult subjects recruited from a single-consultation outpatient attendance with normal biochemistry, represented a sample of our inner city district population. 25-hydroxyvitamin D ($25,OHD_3$) levels were measured to illustrate the effects of season, sex and ethnic group on vitamin D levels and subjected to distribution analysis. This method quantifies as a percentage the distribution of $25,OHD_3$ concentrations (observed concentration, OC) in pooled group data. The data can be expressed as distribution frequency domains or cumulative frequency ogives (0–100%) or transformed into discrete linear probits, amenable to regression analysis. An estimate of the OC_{50} (mid-point) and upper (either OC_{75} or OC_{95}) or lower (either OC_{25} or OC_5) range or at any other frequency between subject groups can be compared.

A marked difference in $25,OHD_3$ levels between Asian and non-Asian asymptomatic adult subjects was seen during both seasons. $25,OHD_3$ deficiency was defined as at or below the OC_{25} for the non-Asian group (for both sexes: winter $<13{\cdot}36$ ng/ml, summer $<13{\cdot}38$ ng/ml). The majority of Asians of both sexes were $25,OHD_3$ deficient (winter 94%, summer 82%).

The distribution analysis provides an easy technique to compare $25,OHD_3$ status of different subject groups, allowing the description of populations using either longitudinal or cross-sectional data. This method may offer a way of describing $25,OHD_3$ deficiency between observers worldwide.

Journal of Endocrinology (2003) **179,** 119–129

Introduction

The diverse role of vitamin D is widely recognized, particularly after discovery of its steroid structure. Vitamin D participates in mineral homeostasis, regulation of gene expression and cell differentiation. There is significant epidemiological, *in vitro* and animal experimental work to suggest that vitamin D deficiency is linked to several clinical conditions (Newmark 1994, Keen *et al.* 1997, Ahonen *et al.* 2000, Cantorna 2000, Hayes 2000, Tokuda *et al.* 2000, Altschuler 2001, Ayesha *et al.* 2001). 25,hydroxyvitamin D ($25,OHD_3$) deficiency has been implicated as a risk factor in syndrome X, diabetes (Boucher 1998) and ischaemic heart disease. Recently McGrath (2001) has proposed that the early origins of these adult problems could arise due to deficiency of vitamin D during 'a critical window of development' in prenatal and perinatal life. For example, during foetal and infant life vitamin D is essential for the central nervous system (Brunvand *et al.* 1996). Thus, vitamin D deficiency during foetal development may cause early imprinting of the functional characteristics of various tissues throughout the body. The impact of this suggests that there is an avoidable risk factor for adult diseases.

Monitoring for vitamin D deficiency, however, still presents a dilemma as no consensus exists between authors worldwide on the definition or biochemical criteria of a vitamin D 'deficiency state'. Cited lower limits based on serum levels of $25,OHD_3$ range between 2 and 18 ng/ml (5–45 nmol/l) (O'Shea & Carter 1998); $25,OHD_3$ in winter <8 ng/ml (20 nmol/l) and summer <10 ng/ml (25 nmol/l) (Bender 1992); or toward levels that will minimize serum parathyroid hormone (PTH) concentrations, i.e. between 20 and 32 ng/ml (50–80 nmol/l) (Vieth 2001).

0022–0795/03/0179–119
Online version via http://www.endocrinology.org

The criteria for replete or insufficient vitamin D would also depend on the clinical impact, and both the level and duration of persistently low $25,OHD_3$ complicate this. For example, although levels of $25,OHD_3$ <8 ng/ml (20 nmol/l) are known to cause clinical symptoms, such low levels will have already been in existence for many months if searching for hypovitaminosis D myopathy (Glerup *et al.* 2000). $25,OHD_3$ levels maintained just below 20 ng/ml (50 nmol/l) can still affect later bone mass (Docio *et al.* 1998). Furthermore, serum PTH measurements can reliably be measured within the normal range only when $25,OHD_3$ levels exceed 40 ng/ml (100 nmol/l) (Brot *et al.* 2001).

The ongoing pathophysiological dilemma of defining deficient vitamin D levels is further complicated when comparing different subject groups. Serum $25,OHD_3$ concentrations are generally measured and compared between small numbers of subjects (usually under 100), with variable characteristics (age, sex, ethnic group, diet, latitude) and are influenced by season. Previously most authors have compared subjects' vitamin D status using mean or median, S.D. and range. However, standard summary statistics utilize parametric methods and assume a normal distribution of measurements, but serum vitamin D levels may often appear as skewed or clustered data. The upper and lower limits are thus less reliably defined. A distribution method allows a visual and graphical display of data enabling comparison of differences in vitamin D throughout the whole data array. This may facilitate clinicians quickly observing relevant patterns between subject groups. (See Appendix for terminology and comparison of statistical methods.)

We illustrate the subject of 25, OHD_3 levels in different ethnic populations (asymptomatic adults) in Birmingham, UK, assayed at different times of year.

We report the use of a suitable method of comparing serum $25,OHD_3$ levels with discussion of the criteria to define deficiency and normal limits appropriate for some populations.

Materials and Methods

The method of 'distributional analysis' is well known and has been used to analyse other hormone data (Matthews *et al.* 1991).

The local research ethics committee approved the proposed investigation for vitamin D data. Briefly, during a 2 year period (1994–1996), 240 adults (median age 50 years, range 17–98; only four subjects were above the age of 75 years) attending for a single-consultation outpatient clinic at a hospital (City Hospital) in Birmingham, with symptoms unrelated to vitamin D status, were included in the study. All were having single-sample blood taken for clinical purposes additional to assay for $25,OHD_3$ levels.

These subjects were non-pregnant asymptomatic adults with normal biochemistry (i.e. normal calcium, phosphate, alkaline phosphatase, renal and liver function).

Subjects were selected purposively ('convenience sample') with stratification to ensure that 30 subjects of each gender and ethnic origin (Asian or non-Asian (Afro-Caribbean and Caucasian)) were included at each of two time-points ('winter' – February/March, and 'summer' – July/August). No subject was included in both seasons. Subject samples were taken over a 2 year period and dated, as such cross-sectional data were obtained from 240 subjects.

Assays

The serum samples were initially centrifuged and stored at -20 °C. Irrespective of season-defined months, samples were analysed in batches within 6 months of sample drawing during the regular Regional Laboratory activity for vitamin D assay needed for clinical purposes (approximately two monthly). The serum $25,OHD_3$ was initially analysed by three assays but only the results of the Incstar RIA kit are reported. The Incstar RIA (^{125}I) kit was obtained from Sorin diagnostics (now known as DiaSorin Medical Systems Group, Saluggia, Italy). This has the advantage of a specific $25,OHD_3$ tracer (although it can accurately measure either D_2 or D_3), excellent specificity and is standardized against HPLC purified 25 OHD_3.

Standard methods to ensure quality control of all specimens were observed (Hollis *et al.* 1993, Bates 1997). Each laboratory run contained labelled coated tubes in duplicate for each standard (total six standards 0–100 ng/ml), control and sample, and a standard curve was prepared for each run.

Specificity Specificity, by estimating the percentage of cross-reaction (50% inhibition for $25,OHD_3$) was 100%; 25 OHD_2 was 0·6%.

Sensitivity Sensitivity analysis suggests that the lowest detectable limit distinguished from the zero standard was 0·6 ng/ml.

Intra-assay variation Within-run variation was determined by replicate determination of two different control sera in one assay: variability coefficient of variation (CV) at 8·6 and 22·7 ng/ml was 9·4 and 8·2% respectively.

Inter-assay variation Between-run variation was determined by replicate measurement of two different control sera in several different assays: variability CV (%) at 33 and 49 ng/ml was 9·1 and 11% respectively.

Accuracy The linearity was checked by serial dilution of two samples of different $25,OHD_3$ concentrations with zero standard solution. Six dilutions (neat to 1:64) were

Table 1 Class intervals or 'domains' used for distribution analysis of serum 25,OHD_3 levels. The 25,OHD_3 concentrations are log-transformed and sorted into ascending order and allocated class intervals that should reflect the range of measured hormone

	Serum 25,OHD_3 (ng/ml)
Class interval	
1	1
2	4
3	8
4	10
5	15
6	20
7	30
8	40
9	50
10	70
11	90

performed for each sample and the diluted samples assayed by the RIA method.

Recovery Normal human serum with known concentration of 25,OHD_3 was enriched ('spiked') with four increasing amounts of 25,OHD_3 (0 to 60 ng/ml) and recovery was between 97·8 and 114%.

External quality control In our laboratory the results of the 25,OHD_3 assay method were consistently within 1 S.D. of the method group mean in the international External Quality Assessment Scheme for this metabolite.

Data analysis

The serum 25,OHD_3 data were first log-transformed for analysis by both parametric and distribution analysis methods. Mathematical transformation does not alter the relationship between each of the original data points, but allows a smoother way of identifying a pattern. Such mathematical transformation is very useful, although not an essential prerequisite when performing distribution analysis. Non-transformed data can be used by most standard methods for large sample sizes (>100 per group).

Distribution analysis *Class intervals* Grouped ranges for the $\log_{10}$ 25,OHD_3 values (Table 1), were generated from all the data, and subsets were then compared using the cumulative distributions across the class intervals.

Groups can be constructed with equal class intervals or equal numbers of observations. The class intervals must be both mutually exclusive and exhaustive. The group intervals do not always have to be equal, but the same intervals should be used for all the datasets to be compared.

Frequency distribution or 'domains' Cumulative frequency distributions were obtained to provide contrasts by gender, ethnicity and season.

Frequency polygon or 'ogive' Calculation of frequency (absolute, relative or cumulative) can be made by hand or by software. Frequency distributions show how many observations on a given variable have a particular attribute. Converting these raw numbers into percentages provides a more useful description of the data. The frequency distribution is a prerequisite for both the various graphs used to display data and the basic statistics to describe a dataset. A frequency polygon (line plot) graph of the cumulative frequency enables characteristics to be easily compared (Kenney & Keeping 1962).

Linear probit transformation The relationship between variables x and y is often 'S' shaped as demonstrated in the cumulative frequency ogive. If the true underlying variable we are predicting is continuous, we can assume the errors are normally distributed and the probit transformation can be used.

Practically, the cumulative distributions of 25,OHD_3 data may either be plotted manually on probability paper or probit values obtained from transform tables (Fisher & Yates 1967). The points used for the x-value of the plot are the mid-group values of the grouped 'cumulative data' from which it is easy to estimate any particular percentile of interest for any subset; and the mid-points of each class interval can be used as points in a regression equation. This provides a statistical method in assessing observed concentration (OC) differences.

The S.D. of the observations can be estimated from the slope of the probit by S.D. = covariance($x.y$)/slope and differences between the lines can be assessed using the method proposed by Schefler (1969).

To calculate a 95% confidence interval for the slope (linear probit) we simply calculated the S.E. of the slope as the equation of the sample regression line is $y=a+bx$ and the coefficient of the variable x is b (the slope). The 95% confidence limits were calculated by the method of Snedecor & Cochran (1989). For each vitamin D level at specified OC parameters, we followed the general form for a confidence interval point estimate ± (critical value) (S.E. of the estimate). Here the point estimate in question is b and its S.E. is given the critical value 2·306 for a 95% confidence interval.

For linear probits, both intercept and slope require careful interpretation. First, the intercept *per se* has little meaning, but horizontal differences between cumulative curves can be used to illustrate differences between estimated percentile points of different subgroups. The assessment as to whether these differences are significant can be made statistically using conventional methods, or clinically, whenever it is a matter of judgement.

Table 2 Observed concentration (OC) parameters of $25,OHD_3$ are compared (A) between sex, ethnic group and season, and (B) with standard statistical analysis. OCs confirm that there is a seasonal variation in both Asians and non-Asians but $25,OHD_3$ levels are significantly lower in Asians in both sexes. Standard statistical analysis shows similar mid-point vitamin D levels but range parameters differ

(A) **OC of $25,OHD_3$ with 95% confidence intervals** (ng/ml)

	Female		Male	
	Asian	Non-Asian	Asian	Non-Asian
Winter OC				
OC_{25}	2·51 (2·40 – 2·62)	12·59 (12·52 – 12·66)	2·14 (2·0 – 2·28)	14·13 (13·99 – 14·27)
OC_{50}	4·57 (4·31 – 4·82)	16·98 (16·76 – 17·20)	4·36 (4·14 – 4·58)	17·37 (17·15 – 17·59)
OC_{75}	8·32 (7·87 – 8·77)	25·12 (24·65 – 25·58)	15·84 (15·53 – 16·15)	28·18 (27·88 – 28·48)
Summer OC				
OC_{25}	6·76 (6·66 – 6·86)	12·30 (12·01 – 12·59)	8·04 (7·87 – 8·21)	14·45 (14·31 – 14·59)
OC_{50}	10·0 (9·82 – 10·18)	19·05 (18·84 – 19·26)	12·30 (12·17 – 12·43)	20·42 (20·20 – 20·64)
OC_{75}	12·88 (12·64 – 13·12)	31·62 (31·33 – 31·91)	16·98 (16·57 – 17·39)	32·36 (32·05 – 32·67)

(B) **Serum $25,OHD_3$ concentration** (ng/ml)

	Sample size	Median	Mean	S.D.	Highest	Lowest
Winter						
Asian female	30	4·95	6·9	7·13	38·2	0·4
Asian male	30	4·85	5·5	2·87	12·5	0·4
Non-Asian female	30	19·55	19·58	7·41	38·1	6·0
Non-Asian male	30	18·80	24·19	8·11	44·5	8·4
Summer						
Asian female	30	9·90	10·53	3·99	21·6	4·7
Asian male	30	13·80	14·27	6·03	34·5	4·9
Non-Asian female	30	20·30	15·96	15·92	88·4	5·6
Non-Asian male	30	25·30	25·38	9·48	47·0	9·9

Secondly, differences between slopes correspond to differences between the S.D.s (more generally, variability) of the various subsets, with a steeper slope representing a smaller S.D. The advantage of this method is that the whole of the data can be compared between groups and the populations compared at any point.

Results

Of the 120 samples collected from each season, 60 were from Asian subjects and 60 from non-Asian subjects, in each case 30 males and 30 females.

The results of standard summary statistics of the vitamin D data are compared within each subgroup with the OC parameters of distribution analysis (Table 2A and B).

Cumulative frequency profile

The performance of this method is illustrated by our $25,OHD_3$ concentrations and compares season, ethnicity and gender. Any given OC parameters can be estimated easily from the cumulative distribution. Figure 1 demonstrates that clear differences in concentration frequency exist throughout the whole range of $25,OHD_3$ measurements.

Effect of ethnic group and season

When comparing OC_{50} or lower (OC_{25}) or upper (OC_{75}) levels of $25,OHD_3$ levels in ethnic groups, the difference is particularly marked in winter.

Asian $25,OHD_3$ levels were significantly lower than non-Asians ($P<0·05$) (Table 2A). In both seasons, over 95% of Asian men and women have a $25,OHD_3$ level below the mid-point (OC_{50}) of their respective sex-matched non-Asian group (Asians vs non-Asian, winter 4·5 vs 17·2, summer 11·2 vs 19·7 ng/ml). The upper level (OC_{75}) was less than half the level seen in non-Asians (Asians vs non-Asian, winter 12·1 vs 26·7, summer 14·9 vs 32 ng/ml).

The lower level (OC_{25}) showed a 6-fold difference between Asians and non-Asians in winter and 2-fold difference in summer (Asians vs non-Asian, winter 2·3 vs 13·4, summer 7·4 vs 13·4 ng/ml).

Probability transformation analysis confirms a significant difference between Asians and non-Asians ($P<0·0025$) during both seasons (Fig. 2).

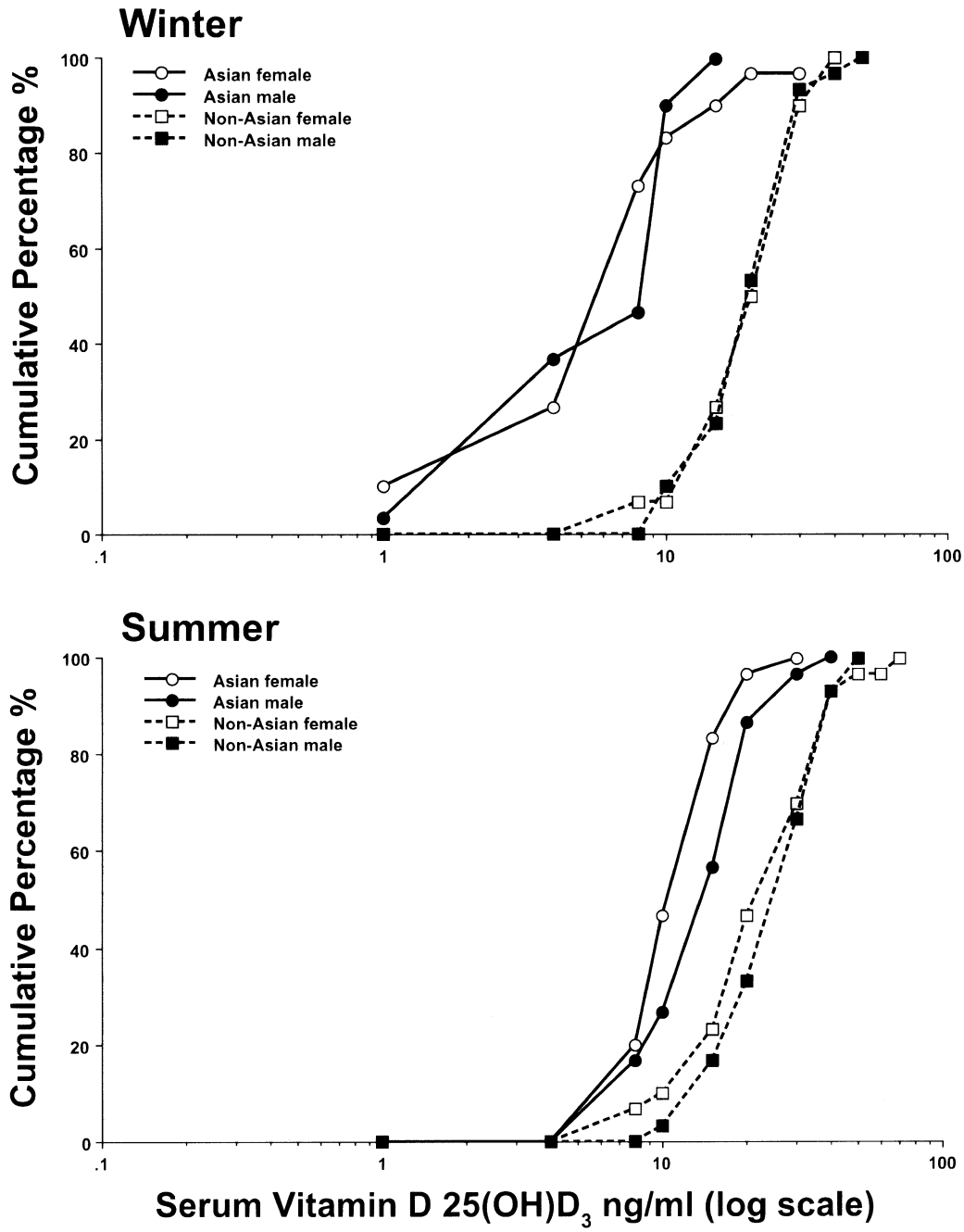

Figure 1 Performance: cumulative distribution of serum 25,OHD_3 levels. The cumulative distribution is a sigmoid curve. These ogives demonstrate visually the differences in serum 25,OHD_3 concentration frequency between two groups: comparing season, sex and ethnic group. An approximate estimate of mid-point and or any comparative parameter is described as 'observed concentrations' (OC parameters) and can be easily made visually from such plots.

Effect of gender

Generally, although Asian women were noted to have the lowest 25,OHD_3 levels in both seasons (Table 2A), the OC_{50} was not statistically different between sexes in either ethnic group (both seasons, female vs male: Asian 7·3 vs 8·3; non-Asian 18·0 vs 18·9 ng/ml). Interestingly, during winter, Asians of both sexes had extremely low OC_{25} 25,OHD_3 levels (male vs female 2·1 vs 2·5 ng/ml).

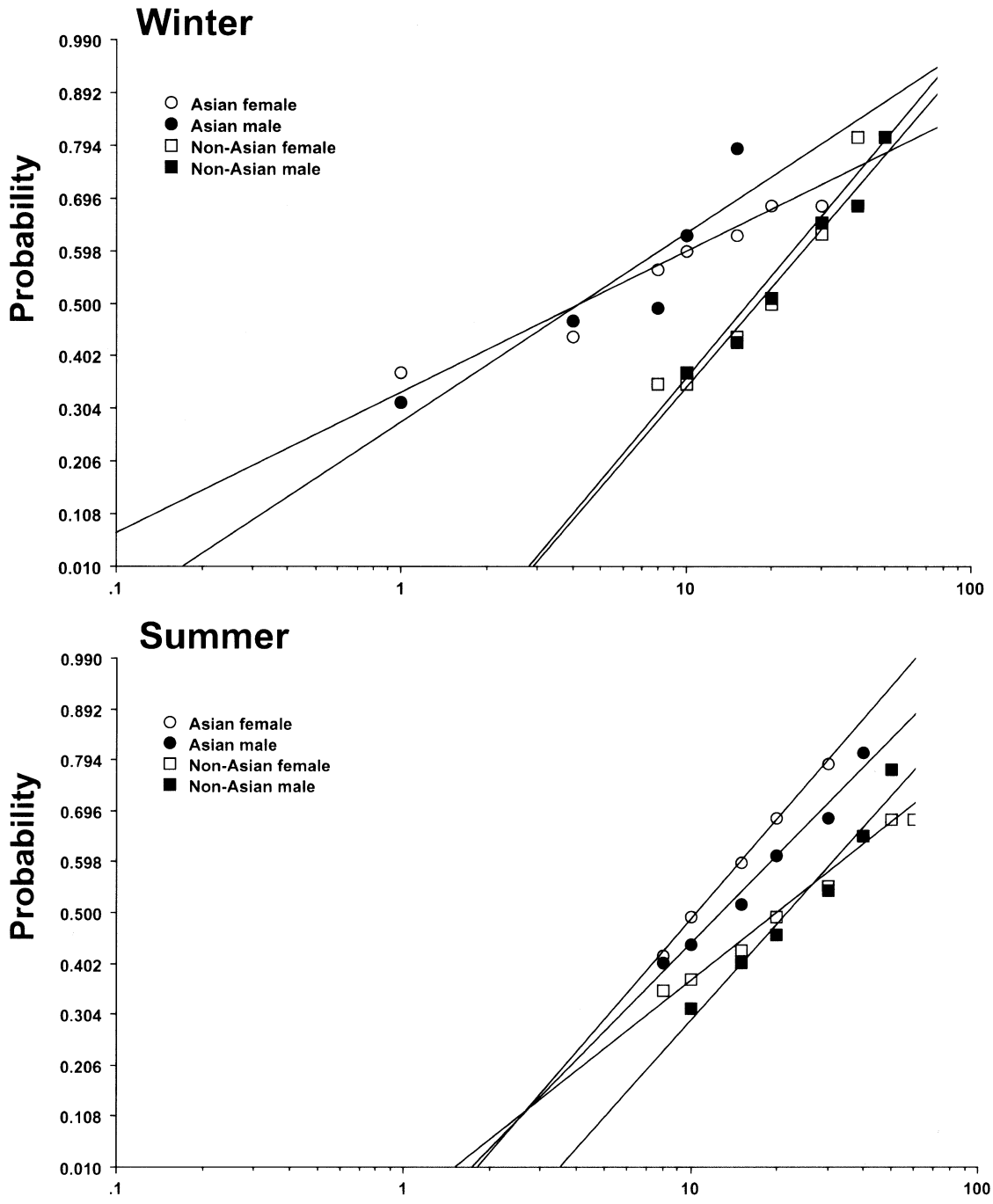

Figure 2 Probability transformation. The Figure displays the lines of regression of the probit values between subject groups comparing sex, ethnic group in each season. These regression lines are significantly different with a null hypothesis probability of $P < 0{\cdot}0025$.

OC_{25} as predictor of deficiency of vitamin D (Table 3)
Clearly the distribution of vitamin D in Asians and non-Asians are very different. It is important to determine if the threshold value of non-Asian OC_{25} enables an accurate prediction of the presence of vitamin D deficiency in both subgroups.

Table 3 Non-Asian OC_{25} parameter is a screening tool to detect vitamin D deficiency in ethnic subgroups: validation. Sensitivity is the fraction of those with vitamin D deficiency correctly identified; specificity is the fraction of those without vitamin D deficiency correctly identified. Postitive predictive value expresses the probability that a person with an OC_{25} level result is deficient; negative predictive value is the probability that a person with a result above OC_{25} is not deficient. Likelihood ratio suggests that an Asian subject with $25OHD_3$ level below non-Asian OC_{25} is 13 times more likely to be deficient than with a level above OC_{25}

	Asians		**Non-Asians**		**Whole group**	
	Negative	Positive	Negative	Positive	Negative	Positive
Vitamin D deficiency OC_{25}=13·8 ng/ml						
Absent	21	4	96	3	117 (a)	7 (c)
Present	3	92	6	15	9 (b)	107 (d)

Sensitivity=d/c+d=107/114=0·938.
Specificity=a/a+b=117/126=0·928.

Predictive value+ve=d/b+d=107/116=0·922.
Predictive value −ve=a/a+c=117/124=0·944.

Likelihood ratio=sensitivity/1 − specificity=13·03.
Prevalence=d+b/a+b+c+d=48% both, 80% Asians, 17·5% non-Asians.

We propose that OC_{25} level for non-Asians is used as a screening tool to detect vitamin D deficiency. Thus, a vitamin D level of 13·8 ng/ml (both seasons, both sexes) is the cut-off level for our subjects. The CV of the assay at 13·8 ng/ml is 9·4%, thus subjects with a vitamin D level of 15·1 ng/ml would be missed (false negative) and subjects with level of 12·5 ng/ml may not be deficient (false positive).

Using the non-Asian OC_{25}, the sensitivity (the fraction of those with vitamin D deficiency correctly identified) for both groups was 93·8%, the specificity (the fraction of those without vitamin D correctly identified) was 92·8%. For the OC_{25} test, the specificity of 92·8% means that 7·2% of subjects who are deficient will test falsely negative (7 of every 100 persons).

The predictive value (PV) depends on both the accuracy of the biochemical test and the prevalence of deficient subjects. As the prevalence is 80% in Asians and 17·5% in non-Asians, with the same test (OC_{25}) the PV changes.

If we test each group of 100 persons by using a test with 92·8% specificity, approximately seven persons in each group will have a false-positive test result.

Non-Asian group The assumed vitamin D deficiency prevalence is 17·5%. Thus, 17·5 true-positive and 7 false-positive test results indicate a PV positive of 71·4% (Table 3).

Asian group The assumed vitamin D prevalence is 80%. With 80 true-positive tests results and 7 false-positive tests results, the PV positive is 92%.

Thus, the PV of the OC_{25} level is very accurate for Asians due to the higher prevalence of vitamin D deficiency.

As the likelihood ratio equals 13, it suggests that a subject with a vitamin D level at non-Asian OC_{25} is 13 times more likely to be deficient than a subject with vitamin D level above OC_{25}. We conclude that the OC_{25} level for non-Asians is an appropriate tool to identify vitamin D deficiency.

Normal range and deficiency

$25,OHD_3$ deficiency depends on season, sex and ethnic group. We could define the $25,OHD_3$ deficiency as either the level of $25,OHD_3$ below OC_{50} or that below OC_{25} for the 'standard population' in each season. The distribution method facilitates a measure of the normal range and defines a 'deficiency' appropriate to the population studied. Any OC parameter can thus be used to compare populations residing at different latitudes.

From our data, the acceptable $25,OHD_3$ levels for our adult group: non-Asians OC_{50} ($OC_{25} - OC_{75}$) is: winter, males 17·4 (14·1–28·2) ng/ml, females 17 (12·6–25·1) ng/ml; summer, males 20·4 (14·5–32·4) ng/ml, females 19·1 (12·3–31·6) ng/ml.

In our population, the definition of $25,OHD_3$ deficiency chosen as the OC_{25} parameter for both sexes, suggests that deficient levels would be <13·4 ng/ml in both seasons. These criteria suggest that 82% of Asians of both sexes have $25,OHD_3$ levels below the OC_{25} of the normal sex-matched population during summer and 94% of Asians are extremely $25,OHD_3$ deficient during winter.

Discussion

The reported prevalence of $25,OHD_3$ deficiency varies widely. In the UK, $25,OHD_3$ deficiency has been reported in Asians: over 50% in community-resident

elderly (Solanki *et al.* 1995); 78% attending a rheumatology clinic (Serhan *et al.* 1999) and 56% in pregnant Asian women (Alfaham *et al.* 1995). We have noted an increased number of Asian infants with hypocalcaemic symptoms (Pal & Shaw 2001). Worldwide, even in abundant sunshine, reports suggest up to 80% of dark-skinned pregnant women in Melbourne exist with subclinical hypovitaminosis D (Grover & Morley 2001, McGrath *et al.* 2001). In New Zealand, significant numbers of young Asian children (3–36 months, median age 12 months) are presenting with extremely low levels of $25,OHD_3$ (Blok *et al.* 2000), and a resurgence of rickets has also been seen in the USA (Rowe 2001, Nesby-O'Dell *et al.* 2002).

The difference in prevalence reported between studies may relate to differences in assays used for $25,OHD_3$, the definition of $25,OHD_3$ deficiency, latitude or other characteristics of the group being studied. Monitoring of $25,OHD_3$ levels in subjects can only be advocated after clarification of a nationally or internationally agreed 'deficient range' for $25,OHD_3$ and the optimum time to screen within the seasonal fluctuations seen over the year. Some authors suggest using a physiological definition of $25,OHD_3$ deficiency (Finkelstein & Thomas 1998) but others exercise caution in the interpretation of these biochemical alterations (Brot *et al.* 2001, Rucker *et al.* 2002).

The distribution method that we propose is a simple way of identifying differences of vitamin D data between subjects. The use of 'observed concentrations' allows the use of any domain (e.g. OC_5, OC_{25}, OC_{50}, OC_{75}, OC_{95}) to be used for comparison by other reporters. The OC measurements reflect the probability that serum $25,OHD_3$ is below a particular level. The distribution percentage enables a display of data (either as 'domain profiles' or as 'sigmoid ogives'). The purpose of graphing is to illustrate, make comparisons easy, avoid distortion and provoke thought about the data.

Cumulative frequency also facilitates conversion into linear probits enabling the differences between groups to be assessed by regression methods (Fox 1997) and confidence intervals. The use of the probit regression model in our example can be justified by positing a random, normally distributed noise process that interferes with the vitamin D pure data by imposing characteristics such as sunlight, season and latitude on the final result. These parameters could equally be considered as a weight function. For simplicity, we have separated these characteristics to describe our data. This method is easy to perform for clinicians, either with or without computer programs and facilitates comparison of vitamin D data between researchers.

Table 2A and B compare the group summary values (standard vs distribution methods). Defining the percentage occupancy of the 'observed concentration' of $25,OHD_3$ at OC_{25} as predicting '$25,OHD_3$ deficiency' enables a comparison of characteristics between subjects in six groups (sample size: total 240, each group of 30). At larger sample sizes (such as during population studies under field conditions, e.g. over 100 per group), other parameters (e.g. OC_5) could be chosen to predict deficiency.

When using standard summary statistics, the central values such as median, mean and OC_{50} are similar (Table 2B) but the distribution values obtained are very different (compare vitamin D values of S.D., highest, lowest and $OC_{25–75}$).

There may be several reasons for this. Although standard statistics are commonly used to summarize data and confirm statistically significant differences between groups, there are often assumptions about the distribution of data. Vitamin D is likely to be geometrically distributed and may be further skewed with small numbers of subjects or if a particular subgroup characteristic exists.

Although the central limit theorem states that as the sample size becomes large, then the sampling distribution of the mean becomes approximately normal regardless of the distribution of the original variable, with smaller sample sizes inaccuracies arise if such normality assumptions are made. Other complex methods could be used to identify the best value for the shape parameter for a range of distributions, but it is simpler to describe the dataset by a distribution method. Frequency distributions can be used to explore each variable in a dataset separately without assumption of its distribution. It looks at the range of values, the central tendency of the values and describes the pattern of response to the variable or describes each variable on its own. The disadvantage is that frequency distribution may not reveal the extreme skew in a dataset, and may de-emphasize the range or extreme values when open classes are used (e.g. over 60 or under 10). The method allows the information from several individuals to be pooled so that analysis of data from groups of individuals with a particular condition or characteristic can be made.

Differences between Asian and non-Asian groups are known to exist (Iqbal *et al.* 1994). We found that in our indigenous non-pregnant adult population a state of subclinical $25,OHD_3$ deficiency exists in a substantially high percentage of an asymptomatic Asian population (winter 94%, summer 82%), when compared with a non-Asian population (OC_{25}) matched for age, sex, latitude and season. Summer sun exposure certainly contributes much to the fluctuations in annual $25,OHD_3$ status in the UK, with higher values between June and October. In healthy middle-aged Danish women (enrolled in a 2·5 year period for the Danish Osteoporosis Prevention Study (Brot *et al.* 2001)) there was a clear relationship between $25,OHD_3$ levels and number of hours of sunshine per month with a lag period of 2 months. The prevalence of $25,OHD_3$ deficiency in the Danish study was 7% rising to a maximum of 32·8% if sunshine was limited. Although this seasonal prevalence is normally seen in healthy subjects, recent concerns of a rising prevalence of wintertime $25,OHD_3$ deficiency is now recognized in the white population (up to 78% of French male adolescents (Guillemant *et al.* 2001), 27·8% healthy females from Southern Italy (Carnevale *et al.* 2001)) and 62% of healthy

non-pregnant women of child-bearing age in Jordan (Mishal 2001). This deficient state can persist despite recommended vitamin D intakes (Vieth *et al.* 2001).

$25,OHD_3$ deficiency was prevalent in our Asian group in both sexes despite the evident seasonal changes. Low $25,OHD_3$ levels existed in the majority (90%) if defined as deficient by both OC_{50} and OC_{25} criteria compared with our non-Asian group. A recent study from Delhi has also found that excessive numbers of healthy Asian subjects of both sexes have low $25,OHD_3$ concentrations whilst residing in abundant sunlight throughout the year (Goswami *et al.* 2000). The extremely high prevalence of $25,OHD_3$ deficiency perhaps suggests a genetic predisposition to $25,OHD_3$ deficiency in Asians.

This suboptimal $25,OHD_3$ status existing in Asians would facilitate the occurrence of symptomatic $25,OHD_3$ deficiency particularly during increased requirements, such as in early childhood, adolescence, pregnancy and during lactation (Islam *et al.* 2002).

Careful attention to maternal vitamin D status may also prevent infant morbidity and impact on a range of adult health outcomes (Hypponen *et al.* 2001, Lamprecht & Lipkin 2001). However, a Cochrane meta-analysis concluded that there is still insufficient evidence to supplement all mothers during pregnancy (Mahomed & Gulmezoglu 2000). Most previous studies regarding use of maternal vitamin D supplementation are too diverse and do not provide the same summary statistic. As only two randomized controlled studies met strict criteria for the Cochrane analysis, the weight of evidence to decide on supplementation should currently be based on professional concern in specific populations or national guidance (Department of Health 1991). The most appropriate dose of vitamin D needed during pregnancy is yet to be defined as animal work has shown that exposure to extremely high amounts of vitamin D during gestation may result in adverse effects on foetal vasculature (Norman *et al.* 2002). Given the high prevalence of $25,OHD_3$ deficiency in Asian women, and the emerging evidence of early origins of adult disease, we would consider it both safe and prudent to provide at least the current recommended supplementation of vitamin D to all Asian women during pregnancy. The immediate and longer-term risks of $25,OHD_3$ deficiency should be identified for all UK Asian families (Shaw & Pal 2002).

Our data confirm that the prevalence of a $25,OHD_3$ deficient state is high, particularly in UK Asians. $25,OHD_3$ deficient levels should be standardized so that reports of prevalence can be compared. We would suggest that the distribution method provides an alternative statistical approach for clinicians when dealing with relatively small numbers (e.g. subgroup characteristics) for comparing either cross-sectional or longitudinal vitamin D data. The method is easy, the whole dataset can be used and an immediate visual description of the data can be generated. It is important to define and detect asymptomatic $25,OHD_3$ deficiency early, as providing adequate vitamin D supplementation may prevent potential long-term consequences.

Acknowledgements

We are grateful to the Department of Biochemistry, City Hospital, Birmingham for providing the raw vitamin D data, Mr D Mayoss-Hurd (Senior Biochemist, City Hospital) for supervising the collection, performance and analysis of all $25,OHD_3$ assays and Mr A Bignall, Head of Biochemistry, City Hospital, Birmingham for his support of this study. We thank Medical Illustration and Graphics Department at Salford Royal Hospitals NHS Trust for reproduction of graphs.

Appendix: Terminology and comparison of statistical methods

Appendix Table 1 Terminology of statistical methods

	Description/advantage
Terminology	
Ogive	Any continuous cumulative frequency curve
Class intervals	Nominal data lists all possible categories, interval data have an appropriate number of data classes
Frequency distributions	a) Condense and summarize large amounts of data in a useful format b) Percentages are reported rather than raw counts to give readers more context c) Describe all variable types d) Facilitate graphic presentation of data e) Begin to identify population characteristics f) Permit comparison of data sets g) Cumulative frequency facilitates conversion into linear probits for probability analysis
Probit	Stands for probability unit
Probit model	$Pr(y=1x)=\Phi(xb)$ Pr is the inverse of the cumulative standard normal distribution function, often referred as probit or normit and Φ is the standard cumulative normal probability distribution and xb is the probit score (or index)

Appendix Table 2 Comparison of statistical methods

	Distribution assumption	Sample size preferred	Other comment	Limitation
Statistical test				
t-test (two sample)	Normal	Large >40	Testing average or mean; small datasets with substantial non-normal distribution can mask real difference	If groups differ in another characteristic, apart from mean
Mann–Whitney/ Wilcoxon rank-sum	Non-parametric	Independent random samples	Combining both samples into one set, median values compared	Assume distribution same in both groups
Kolmorgorov–Smirnov	Continuous specified	Unequal sets acceptable	More sensitive at centre of distribution than at tails	Location, scale and shape must be determined
Distribution analysis	Any type all data included	Any size large or small	Inclusive of all data	Ensure class intervals are exhaustive

References

Ahonen MH, Tenkanen L, Teppo L, Hakama M & Tuohimaa P 2000 Prostate cancer risk and prediagnostic serum 25-hydroxyvitamin D levels (Finland). *Cancer Causes and Control* **11** 847–852.

Alfaham M, Woodhead S, Pask G & Davies D 1995 Vitamin D deficiency: a concern in pregnant Asian women. *British Journal of Nutrition* **73** 881–887.

Altschuler EL 2001 Low maternal vitamin D and schizophrenia in offspring. *Lancet* **358** 1464.

Ayesha I, Bala TS, Reddy CV & Raghuramulu N 2001 Vitamin D deficiency reduces insulin secretion and turnover in rats. *Diabetes, Nutrition and Metabolism* **14** 78–84.

Bates CJ 1997 Vitamin analysis. *Annals of Clinical Biochemistry* **34** 599–626.

Bender DA 1992 *Nutritional Biochemistry of the Vitamins* edn 2. pp 77–108. Eds. Cambridge, New York: Cambridge University Press.

Blok BH, Grant CC, McNeil AR & Reid IR 2000 Characteristics of children with florid vitamin D deficient rickets in the Auckland region in 1998. *New Zealand Medical Journal* **113** 374–376.

Boucher BJ 1998 Inadequate vitamin D status: does it contribute to the disorders comprising syndrome 'X'? *British Journal of Nutrition* **79** 315–327.

Brot C, Vestergaard P, Kolthoff N, Gram J, Hermann AP & Sorensen OH 2001 Vitamin D status and its adequacy in healthy Danish perimenopausal women: relationships to dietary intake, sun exposure and serum parathyroid hormone. *British Journal of Nutrition* **86** S97–S103.

Brunvand L, Quigstad E, Urdal P & Hang E 1996 Vitamin D deficiency and fetal growth *Early Human Development* **45** 27–33.

Cantorna MT 2000 Vitamin D and autoimmunity: is vitamin D status an environmental factor affecting autoimmune disease prevalence? *Proceedings of the Society for Experimental Biology and Medicine* **223** 230–233.

Carnevale V, Modoni S, Pilieri M, Di Giorgio A, Chiodini I, Minisola S, Vieth R & Scillitani A 2001 Longitudinal evaluation of vitamin D status in healthy subjects from southern Italy: seasonal and gender differences. *Osteoporosis International* **12** 1026–1030.

Department of Healt 1991 Dietary reference values for food energy and nutrients for the United Kingdom. *Report on Health and Social Subjects 41*. London: HMSO.

Docio S, Riancho JA, Perez A, Olmos JM, Amado JA & Gonzalez-Macias J 1998 Seasonal deficiency of vitamin D in children: a potential target for osteoporosis-preventing strategies? *Journal of Bone and Mineral Research* **13** 544–548.

Finkelstein JS & Thomas MK 1998 Hypovitaminosis D in medical inpatients. *New England Journal of Medicine* [Letter] **339** 346.

Fisher RA & Yates F 1967 *Statistical Tables for Biological, Agricultural and Medical Research*, several editions. Edinburgh: Oliver and Boyd.

Fox J 1997 *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage Publications Inc.

Glerup H, Mikkelsen K, Poulsen L, Hass E, Overbeck S, Andersen H, Charles P & Eriksen EF 2000 Hypovitaminosis D myopathy without biochemical signs of osteomalacic bone involvement. *Calcified Tissue International* **66** 419–424.

Goswami R, Gupta N, Goqwami D, Marwaha RK, Tandon N & Kochupillai 2000 Prevalence and significance of low 25-hydroxyvitamin D concentrations in healthy subjects in Delhi *American Journal of Clinical Nutrition* **72** 472–475.

Grover SR & Morley R 2001 Vitamin D deficiency in veiled or dark-skinned pregnant women. *Medical Journal of Australia* **175** 251–252.

Guillemant J, Le HT, Maria A, Allemandou A, Peres G & Guillemant S 2001 Wintertime vitamin D deficiency in male adolescents: effect on parathyroid function and response to vitamin D3 supplements. *Osteoporosis International* **12** 875–879.

Hayes CE 2000 Vitamin D: a natural inhibitor of multiple sclerosis. *Proceedings of the Nutrition Society* **59** 531–535.

Hollis BW, Kamerud JQ, Selvaag JQ, Lorenz JD & Napoli JL 1993 Determination of vitamin D status by radioimmunoassay with an ^{125}I-labelled tracer. *Clinical Chemistry* **39** 529–533.

Hypponen E, Laara E, Reunanen A, Jarvelin MR & Virtanen SM 2001 Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. *Lancet* **358** 1500–1503.

Iqbal SJ, Kaddam I, Wassif W, Nishol F & Walls J 1994 Continuing clinically severe vitamin D deficiency in Asians in the UK (Leicester). *Postgraduate Medical Journal* **70** 708–714.

Islam MZ, Lamberg-Allardt C, Karkkainen M, Outila T, Salamatullah Q & Shamim AA 2002 Vitamin D deficiency: a concern in premenopausal Bangladeshi women of two socio-economic groups in rural and urban region. *European Journal of Clinical Nutrition* **56** 51–56.

Keen RW, Egger P, Fall C, Major PJ, Lanchbury JS, Spector TD & Cooper C 1997 Polymorphisms of the vitamin D receptor, infant growth, and adult bone mass. *Calcified Tissue International* **60** 233–235.

Kenney JF & Keeping ES 1962 Ogive curves. In *Mathematics of Statistics*, part 1, §2·7, edn 3, pp 29–31. Princeton, NJ: Van Nostrand.

Lamprecht SA & Lipkin M 2001 Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis. *Annals of the New York Academy of Sciences* **952** 73–87.

Mahomed K & Gulmezoglu AM 2000 Vitamin D supplementation in pregnancy. *Cochrane Database of Systematic Reviews* 2:CD000228.

Matthews DR, Hindmarsh PC, Pringle PJ & Brook CG 1991 A distribution method for analysing the baseline of pulsatile endocrine signals as exemplified by 24-hour growth hormone profiles. *Clinical Endocrinology* **35** 245–252.

McGrath J 2001 Does 'imprinting' with low prenatal vitamin D contribute to the risk of various adult disorders? *Medical Hypotheses* **56** 367–371.

McGrath JJ, Kimlin MG, Saha S, Eyles DW & Parisi AV 2001 Vitamin D insufficiency in south-east Queensland. *Medical Journal of Australia* **174** 150–151.

Mishal AA 2001 Effect of different dress styles on vitamin D levels in healthy young Jordanian women. *Osteoporosis International* **12** 931–935.

Nesby-O'Dell S, Scanlon KS, Cogswell ME, Gillespie C, Hollis BW, Looker AC, Allen C, Doughertly C, Gunter EW & Bowman BA 2002 Hypovitaminosis D prevalence and determinants among African American and white women of reproductive age: third National Health and Nutrition Examination Survey, 1988–1994. *American Journal of Clinical Nutrition* **76** 187–192.

Newmark HL 1994 Vitamin D adequacy: a possible relationship to breast cancer. *Advances in Experimental Medicine and Biology* **364** 109–114.

Norman P, Moss I, Sian M, Gosling M & Powell J 2002 Maternal and postnatal vitamin D ingestion influences rat aortic structure, function and elastin content. *Cardiovascular Research* **55** 369–374.

O'Shea D & Carter GD 1998 Hypovitaminosis D in medical inpatients. *New England Journal of Medicine* **339** 345–346.

Pal BR & Shaw NJ 2001 Rickets resurgence in the United Kingdom: improving antenatal management in Asians. *Journal of Pediatrics* **139** 337–338.

Rowe PM 2001 Why is rickets resurgent in the USA? *Lancet* **357** 1100.

Rucker D, Allan JA, Fick GH & Hanley DA 2002 Vitamin D insufficiency in a population of healthy western Canadians. *Canadian Medical Association Journal* **166** 1517–1524.

Schefler WC 1969 *Statistics for the Biological Sciences*, edn 1. Reading, MA: Addison-Wesley.

Serhan E, Newton P, Ali HA, Walford S & Singh BM 1999 Prevalence of hypovitaminosis D in Indo-Asian patients attending a rheumatology clinic. *Bone* **25** 609–611.

Shaw NJ & Pal BR 2002 Vitamin D deficiency in UK Asian families: activating a new concern. *Archives of Disease in Childhood* **86** 147–149.

Snedecor GW & Cochran WG 1989 *Statistical Methods,* edn 8. Ames, IA: Iowa State University Press.

Solanki T, Hyatt RH, Kemm JR, Hughes EA & Cowan RA 1995 Are elderly Asians in Britain at a high risk of vitamin D deficiency and osteomalacia? *Age and Ageing* **24** 103–107.

Tokuda N, Kano M, Meiri H, Nomoto K & Naito S 2000 Calcitriol therapy modulates the cellular immune responses in hemodialysis patients. *American Journal of Nephrology* **20** 129–137.

Vieth R 2001 Would prehistoric human 25-hydroxyvitamin D concentrations be beneficial, and how much vitamin D do we need to ensure desirable nutritional targets? In *Nutritional Aspects of Osteoporosis,* pp 173–195. Eds P Burckhardt, R Heaney & B Dawson-Hughes. San Diego, CA: Academic Press.

Vieth R, Cole DE, Hawker GA, Trang HM & Rubin LA 2001 Wintertime vitamin D insufficiency is common in young Canadian women and their vitamin D intake does not prevent it. *European Journal of Clinical Nutrition* **55** 1091–1097.

Received in final form 26 June 2003
Accepted 17 July 2003